

Quality assurance and disciplinary differences

Tony Becher
University of Sussex

Introduction

My main purpose in choosing this theme is to illustrate what I would argue to be a quite general tendency: namely for university policies and procedures to be framed with little regard for significant contrasts between the characteristic features of different academic disciplines. To this end, I will explore how some of the now quite common approaches to establishing academic quality impinge on aspects of research, undergraduate teaching and postgraduate work in a variety of disciplinary contexts. The paper will conclude by offering some brief speculations about the apparent neglect of subject differentiation in current approaches to quality assurance.

As successive demands have been imposed on higher education institutions by government in Australia, England and elsewhere for quality maintenance and enhancement, it has become more and more apparent that quality is itself a highly elusive concept, giving rise to a range of interpretations and generating a diversity of approaches to its assessment (Lindsay 1993). Much critical comment has focused on the associated notion of 'performance indicators' (see, for example, Cave et al 1988; Goedegebuure 1990; Johns and Taylor 1990), though this has apparently had only a modest effect on the continuing fascination of politicians with quantitative measurement techniques. However, as the following discussion will attempt to show, even some of the more qualitative attempts to frame judgements of merit and worth are open to the charge of operating inequitably between one disciplinary field and another.

Some quality criteria for research

Since research has been subjected to quality assessment for longer than teaching - indeed, it could be argued that research activity has always been closely linked to peer group judgement (Becher 1989) - it may seem appropriate to begin by considering three well-established criteria for excellence in research. All three have been selected, incidentally, to form the basis of the periodic research rating exercises undertaken in England, first by the University Grants Committee, next by the Universities Funding Council and now by the Higher Education Funding Council (England). The criteria in question are the number of publications produced by a particular subject group in each university in a given timespan, the amount of external funding attracted and the coherence of the forward research plans of the department or other unit in question.

Chemists are well favoured in the publication context, since their papers are typically brief (4000 words tends to be on the high side), and commonly multi-authored. An active researcher can expect to produce ten or a dozen such contributions a year: the most prolific individuals can notch up totals of four to five hundred titles over a working life. But where chemists' research problems are commonly small-scale and sharply-defined, and their methods capable of straightforward replication from one context to another, high energy physicists tend to involve themselves with large-scale long-term experiments whose publication yield is relatively small, with the resulting honours often having to be divided between fifty contributors or more. In the very different intellectual arena occupied by historians a productive academic might average one 8000-12000

word article per annum, and perhaps a book every four or five years. Counting titles across these diverse fields therefore offers an unsure guide, even to research productivity, let alone to the quality of the work done, though it gives a more relevant indication of the nature and scale of the characteristic research problems in a particular discipline.

External funding is in its turn a fair measure of the dependence of a research field on large teams and expensive apparatus. Here subjects such as astrophysics and radioastronomy figure prominently, though many other specialisms in the pure and applied sciences are also in a position to make a reasonable case for special grants in respect of equipment, materials and support staff. Those engaged in social science research are less well favoured, since their work seldom involves instrumentation and their teams (if they do not work solo) usually number no more than three or four people. Such costs as are directly attributable to research typically relate to travel and subsistence, plus perhaps an element of computer time. For academics in the humanities, even such potential claims on grant funds may not be justified, so that they necessarily make a poor showing in the grant-getting stakes. The absence of funding is, however, no indication that their work is of little value, even though that criterion can be made to apply in a rough-and-ready way to scientific enquiry. As in the case of publication rates, the measure is most directly indicative of the nature of the research activity in question rather than of its degree of excellence.

At first glance, the requirement for a basic unit to give a coherent account of its future plans for research seems reasonable enough. Certainly, where sizeable numbers of academics are collectively engaged on some particular area of enquiry, and especially where their work depends on expensive gadgetry and consumables, there may be good grounds for asking them to produce some assurance that continuing expenditure at a high level is justified, that they are working towards some broadly identifiable research goal and that they have some grounds for expecting to achieve it in the foreseeable future. It may even be justifiable to expect different research groups within the same department to be engaged in mutually compatible and perhaps neatly complementary topics. Many scientific activities are subject to a contextual imperative, where the structure of knowledge is close-knit, sequential and progressive, where there is a discernable research frontier, and where the emergent research problems are often readily identifiable (Becher 1989).

In the humanities and social sciences, however, it is more appropriate to talk of contextual association, which relates to a looser, reiterative knowledge structure, in which it is not possible to identify a moving frontier of research and where new problems are not sharply determined by the state of existing knowledge - where, indeed, previous topics may readily be tackled afresh. Enquiries in fields characterised by contextual association are understandably more individualistic, more widely dispersed and more disjointed from one another than those in areas governed by a contextual imperative. So in departments of history, literature, sociology, anthropology and the like, there is very rarely any coherent pattern to the research in progress, and still less, any reasonable prospect of

discerning a tidy plan for the years ahead. Where such departments are called upon to demonstrate their research quality in terms of forecasts of their future activity, the exercise has to be one of creative fiction, in which what is to be judged is the ability to construct plausible fantasies rather than to describe the untidy reality. Certainly, the criterion in non-scientific subjects at least can be seen as bizarre and inappropriate, and the question whether the projected research is up to standard is neither here nor there.

It may be noted, in relation to these criteria, that not only do the first two overtly employ considerations of quantity as surrogates for those of quality, but that all three also superimpose criteria more or less appropriate to the sciences (the latter two, but not the first, particularly favouring 'big science') on disciplines in the social sciences and humanities to which they are largely irrelevant. It is as if those concerned with academic policy-making are dominated by a stereotype of academic research which equates all types of intellectual enquiry with its most highly visible manifestations in such areas as nuclear physics and medical biochemistry.

The defence which may be offered against these objections is that, at least as far as the English research selectivity exercise is concerned, the criteria are not applied comparatively across disciplines but are only used to rate research quality separately within each subject field. Although that is true, it does not provide a justification for using procedures which are, in a large number of fields, unreliable indicators of quality for the reasons already noted. Moreover, although the process of reaching quality judgements is subject-specific at the level of the system as a whole, invidious cross-disciplinary comparisons on the basis of these judgements are certainly made at the level of the individual institution. All in all, it has to be said that criteria relating to publication rates, levels of grant funding and coherence of forward research plans are inequitable across the range of disciplinary areas and that their adoption is by that token seriously flawed.

The assessment of teaching

So far we have been concerned with approaches to quality assessment whose weakness lies in their failure to recognise the intrinsic differences between disciplines arising from their epistemological characteristics. In reviewing some common measures applied to the quality of undergraduate teaching, we may observe the similarly inequitable consequences of ignoring extrinsic - that is broadly social - considerations. The three criteria to be considered in this context are the attractiveness of a given degree programme in terms of the ratio of applicants to places, the staff-student ratio in the department concerned, and the employability of its graduates. As it happens, these are among the criteria put forward by the Higher Education Funding Council (England) for rating course provision as excellent, satisfactory or unsatisfactory.

At a commonsense level, one might expect the popularity of a given department's course offerings to reflect the quality of the teaching associated with them. On closer consideration, however, the mechanisms of the market can be seen to operate as inadequately here as they do elsewhere in higher education. In the first place, some subjects enjoy an advantage that others do not. In England at least, there is a serious shortage of well-qualified science teachers in schools (science graduates can earn considerably more by going into industry or by taking up specialist posts in the public services). As a result, science is often taught badly, and students are put off from taking higher education courses in subjects such as physics and engineering. Against this, schools are usually able to recruit the cream of history and English graduates, and thus to achieve a high standard of teaching. This is reflected in the very high proportion of students applying for university places in the same and related subjects. Here, then, is an instance in which the criterion systematically discriminates against the sciences in favour of the humanities. Again, it may be said in extenuation that at the national level each subject is given separate consideration, so that there are no invidious comparisons. The riposte is the same - that such comparisons are

nonetheless drawn between one department and another within the universities themselves. But there are many other distorting factors as well. Students' choices of course and university are in many cases only very tenuously related - if they are related at all - to the perceived quality of teaching. Relevant factors include currently fashionable preferences for particular subjects (media studies and business studies are at present enjoying a vogue in England) or for particular institutions; the amenities - including affordable accommodation - offered in some geographical locations as against others; the nearby presence of friends or relatives; the recommendations (usually based on dated knowledge) of respected teachers in the school; and so on (see, for example, Evans 1988, 1993). It may be that courses which are taught outstandingly badly begin over time to acquire a negative reputation, and that this is eventually reflected in poor recruitment. The converse, however - that courses which attract large numbers of applicants do so because of the excellence of their teaching - is not a tenable proposition: and hence the criterion based upon it must be considered invalid.

Staff-student ratios are an old favourite in the quality measurement stakes, since they are relatively easily calculated and everyone knows what they signify. However, some ambiguity has crept into their interpretation with the heavy emphasis now placed by politicians on cost-effectiveness. There was a time when departments which had relatively few students per member of staff - 1:10 was a favoured norm - were considered likely to provide a higher standard of teaching than those with a less advantageous balance between the two. However, virtue is currently ascribed to more economical arrangements. A ratio of 1:10 would be considered grossly wasteful, while double that would not be seen as out of the way. But even leaving aside the question of interpretation - what ratio is properly indicative of excellence without extravagance? - a problem remains which is closely connected with the criterion of popularity among applicants. If student intake numbers fall to any significant degree over a short timespan, there is bound to be a lowering in the numbers of students per member of staff, just as a rapid increase in recruitment must yield a less favourable ratio, since compensating adjustments in staff numbers cannot be made overnight. This criterion, alongside the percentage of applicants to places, is in actuality a function of several extraneous factors. It is to some significant extent a reflection of the intake pattern, and hence not an independent variable. Moreover, when staff-student ratios are considered in relation to teaching quality, allowance has to be made for the modes of teaching characteristic of different disciplines. Some forms of laboratory work, for example, call for a substantial staff presence, as do seminar and tutorial requirements in subjects such as philosophy and anthropology, while in other fields - law and mathematics would be cases in point - there is an established tradition of lecturing to large groups. Here, as elsewhere, account has to be taken of legitimate variations between one discipline and another.

The third measure of teaching quality, the employability of graduates, is open to comparable objections. As a feature of academic output, it discriminates in the opposite direction to the criterion relating to input (the applicants to places ratio), in that it has a built-in bias in favour of physics and engineering and against historical and literary studies. That bias has nothing to do with the quality of teaching in the subjects in question, being rather a reflection of the current state of the labour market. Before the recession, there were more vacancies on offer to graduates in the pure and applied sciences than there were job opportunities for students in the humanities and social sciences, so that in the short term allowed by the Funding Council - one year from graduation - the latter came off badly in comparison with the former. At present, the criterion is virtually useless, since employment prospects generally are so poor. Its absurdity is brought out by the fact that the percentage of graduates employed within a year has been drastically reduced even in a subject such as electrical engineering: if the test were literally applied, it would signal a sudden and dramatic reduction in the teaching standards of the departments concerned.

Quite apart from this, there are geographical factors which operate to undermine comparability even between like departments, such as the local presence or otherwise of major international companies and the incidence in the region of small enterprises with limited numbers of vacancies at any given time. Here again, it seems on closer analysis that a criterion purporting to assess the excellence of teaching programmes is a better measure of something else - in this case, the state of the market for graduate employment.

Appraising postgraduate programmes

So far, the criteria which have been considered have been predominantly quantitative, based on data capable of being expressed as ratios. Readers may consider that these are easy game for critical comment, given the fundamental absurdity of trying to gauge quality in quantitative terms. While that may indeed be the case, such criteria are deployed in practice by governmental agencies concerned with quality assessment, so their inadequacies deserve to be exposed. There are also instances, however, in which qualitative considerations predominate. The appraisal of graduate programmes is a case in point. We may in this context briefly consider three sets of considerations related to the excellence or otherwise of such programmes, namely the length and intensity of formal research training, the extent of pre-thesis publication and the nature of the doctoral thesis. The discussion which follows will draw extensively on two recent studies of postgraduate education (Becher, Kogan and Henkel 1993; Clark 1993).

It has been argued (ESRC undated) that the quality of postgraduate work is closely dependent on the availability and extent of systematic training provision. This contention, reasonable though it may appear, tends to elicit quite varied reactions from different disciplinary groups. The economists are most strongly in favour of the introduction of a full year's formal postgraduate teaching *en route* to the doctorate. On closer investigation, they see such teaching as a means of bridging the gap between work at undergraduate level and the more highly specialised and mathematically sophisticated requirements of a PhD. Physicists are more equivocal: they accept the need for a measure of specialised preparation - occupying, say, a couple of terms rather than a full year - but insist that this should be localised and specific to the laboratory to which the doctoral candidate is attached, rather than provided across the department as a whole. Among sociology departments, there is a general readiness to comply with Research Council requirements, but the courses provided tend to focus on areas - sociological theory and quantitative methods - which turn out to be irrelevant to the needs of most students. In biochemistry and history the proposal to introduce collective formal training is met with fierce resistance, on the grounds that in both disciplines every research topic is highly individualistic, requiring its own specific training. The contention is that any given course would meet the needs of only a small proportion of doctoral candidates, and would not therefore be justified: instead, ways have to be found of providing for each student's needs as they arise.

These very different responses to the incidence of formal training as a guide to quality in doctoral work provide an interesting reflection of the nature of the disciplines concerned. The lack of uniformity makes it extremely difficult to apply the criterion in question, except by arguing that all history and biochemistry doctorates are of poor quality and all economics ones are excellent - a contention which seems as unjust as it is implausible.

Another criterion which has been put forward in connection with the appraisal of postgraduate work is the quality and extent of pre-thesis publication. We have already noted some of the pitfalls involved in using numbers of publications to assess the value of research. In this context, however, there are added complications arising from differences in disciplinary practice. The economists are once again enthusiastically in favour of encouraging doctoral students to publish journal articles en route to their thesis, for a reason which will be explained shortly. In biochemistry, research students

are seen as members of a laboratory production team, whose publishable writings are closely vetted by their supervisors and are attributed jointly to student and supervisor. In this context, it is in everybody's interest for postgraduates to publish as frequently as possible. Historians are much more ambivalent, partly on the grounds that a substantial piece of historical research is not amenable to being chopped up into article-sized segments and partly because writing a thesis is seen as a highly demanding and time-consuming activity which leaves little scope for the distraction of preparing papers for journal publication. Again, this diversity of disciplinary practice makes it highly questionable to relate the publications of doctoral candidates to the quality of their programmes.

Even what seems to be an eminently sensible criterion - namely the nature and extent of the doctoral thesis itself - is subject to similar problems. As it turns out, not all these are alike in their structure and their expected word length. In a number of major universities, the economics departments have succeeded in negotiating a special arrangement whereby theses in some (but not all) specialisms can take the form of three more-or-less loosely connected papers of article length and publication standard (hence their predilection for candidates to publish materials in advance of submission). The argument behind this procedure is that articles rather than books are now the favoured form of publication in most areas of economics. A three-paper format of this kind does of course bear little relation to the standard 80,000 - 100,000 word dissertation called for in most other social science disciplines and in all humanities subjects. The situation is complicated still further by some of the more highly-rated theses in mathematics and theoretical physics, where the elegance of the solution to the problem initially posed may result in a relatively brief offering of 100 pages or less.

The failure of even these qualitatively-based criteria to distinguish good doctoral work from bad is once more a consequence of the lack of homogeneity of the different disciplines concerned. It seems evident from these and the earlier examples relating to the evaluation of research and the assessment of teaching that attempts to develop criteria which override disciplinary differences are questionable, to say the least.

A blind spot in quality management

It does not follow from the arguments which have so far been advanced that a concern to monitor and promote quality in higher education is pointless or incapable of realisation. What does emerge is that each disciplinary community - and even to some extent each specialist subgroup within that community - has its own particular way of going about its work. The significant cultural differences which can be identified between disciplines rule out as inadequate and misleading the imposition of standard, across-the-board measures of departmental performance. But despite the often quite fundamental diversities between the values and practices of one academic group and those of even an apparently close intellectual neighbour, they are all subject to the same broad considerations. Each can reasonably be asked to give an account of the quality of its research, its undergraduate teaching and its postgraduate provision.

What it is important to recognise is that, while the nature of these requirements may be common across the institution as a whole, the ways in which they can best be met need to be seen as capable of modification in relation to the discipline under review (Becher 1992). That is to say, even if an overall framework for quality assurance is seen as desirable, that framework should within reason allow for justifiable idiosyncrasies if it is to be responsive to the intrinsic and extrinsic variations in particular disciplinary settings.

There remains the question why academic managers, despite the evident inadequacy of the types of quality assessment procedure reviewed in this paper, persist in imposing them on a system which is often hard put to respond - or which, worse still, may have to distort the evidence in order to do itself justice. Why do apparently intelligent people, many of whom know that system from the inside,

put forward such ill-judged measures for assessing the quality of academic activity? There could be many possible explanations: I will put forward four hypotheses which occur to me, without claiming that they are either exhaustive or mutually exclusive.

First, it may be argued that academic administrators, in their concern to meet political requirements for accountability, suppress their own disciplinary affiliations. Like expatriates, they cut themselves off from their native community and become out of touch with its characteristic way of life. In doing so, they lose the sense of how diverse the range of disciplines is with which they have to deal, and begin to view the university world as homogeneous.

The second hypothesis rests on the principle of inertia. It is troublesome to have to manage a system which calls for numerous exceptions to every common procedure. Life for the academic administrator is a great deal easier if all arrangements can be kept at a level of comfortable uniformity. So those concerned with quality assurance in higher education have a strong incentive to turn a blind eye to disciplinary differences and to persuade themselves that standardised procedures are adequate for their purpose.

The next explanation relates to the basic human need to rationalise, ascribing a sense of order to what look like messy phenomena. Graham Allison (1971) portrayed this tendency in a convincing way in his powerful analysis of the Cuban missile crisis. As he showed, the actions of the US Government could be convincingly explained in terms of the influence of bureaucratic rivalries between the State Department and the Department of Defense, and also with reference to the micropolitical behaviour of Kennedy's inner circle of advisers. Nevertheless, a strong tendency remained to adopt what he labelled as the 'rational actor' model, which involved the ascription of highly rational behaviour to Kennedy and Krushchev as the key individuals involved. Similarly, it might be suggested that although the behaviour of different academic groups fails to conform to any neatly consistent pattern, there is an inevitable tendency to impose a sense of order upon it and to represent it as a respectably tidy field of activity.

It is however my fourth and final hypothesis which I find the most persuasive, in that it is part of a general tendency which characterises policy-making in higher education in both England and Australia. Good administration is, by common consent, about fairness - about treating everyone alike. Accordingly, all departments must be subject to the same regimes and regulations: there can be no allowance made for 'special pleading'. If this equity principle results, on a number of occasions, in rough justice, it is at least

justice of a kind. Even-handedness cannot allow for exceptions, since to make them would by definition fail to be evenhanded.

Such a stance may seem entirely defensible to those concerned with managing the higher education system, and by extension to those charged with ensuring the quality of the processes within that system. The academics who suffer the consequences, however, may understandably view it as profoundly mistaken. I have attempted to illustrate a number of sources of bias in relation to indicators of quality, which are no more acceptable because, as Johnes and Taylor (1990) have shown, they sometimes cancel one another out.

The main implication of the argument which has been advanced in this paper should be evident enough. It is that being a good academic manager must involve, among other things, a sensitivity to the varied characteristics and needs of different departmental groups, and a readiness to adapt and modify procedures to take those characteristics and needs properly into account. Quality assurance provides only one particular case of this general proposition.

References

- Allison, Graham T. (1971), *Essence of decision*, New York, Little, Brown.
- Becher, Tony (1989) *Academic tribes and territories*, Milton Keynes, Open University Press.
- Becher, Tony (1992) 'Making audit acceptable: a collegial approach to quality assurance', *Higher Education Quarterly* 46, 1, 47-66.
- Becher, Tony, Kogan, Maurice and Henkel, Mary (1993) *Graduate education in Britain*, London, Jessica Kingsley.
- Cave, M., Haney, S., Kogan, M. and Trevett, G. (1988) *The use of performance indicators in higher education*, London, Jessica Kingsley.
- Clark, Burton R. (ed) (1993) *The research foundations of graduate education*, Berkeley, University of California Press.
- Economic and Social Research Council (undated), *Postgraduate training*, London, ESRC.
- Evans, Colin (1988) *Language People*, Milton Keynes, Open University Press.
- Evans, Colin (1993) *English people*, Buckingham, Open University Press.
- Goedegebuure, L.C.J., Maassen, P.A.M. and Westerheijden, D.J. (1990) *Peer review and performance indicators*, Utrecht, VITgeverij Lemma.
- Johnes, Joanna and Taylor, Jim (1990) *Performance indicators in higher education*, Buckingham, Open University Press.
- Lindsay, Alan (1993) 'Performance and quality in higher education', *Australian Universities' Review* 36, 1, 32-35